R E P O R T R E S U M E S

ED 013 019 FL 000 118

EXAMINATIONS--THEIR USE AND ABUSE.

BY- BENNETT, PEARL S.

PUB DATE 66

EDRS PRICE MF-$0.25 HC-$0.52 13P.

DESCRIPTORS- *EDUCATIONAL THEORIES, *MULTIPLE CHOICE TESTS, *TEST VALIDITY, *TESTING PROBLEMS, *LANGUAGE TESTS, SECOND LANGUAGE LEARNING, ESSAY TESTS, RUSSIAN, STANDARDIZED TESTS, MEASUREMENT GOALS, TEST CONSTRUCTION, ACHIEVEMENT TESTS,

ALTHOUGH MOST EDUCATORS AGREE THAT EXAMINATIONS PERFORM AN IMPORTANT FUNCTION IN APPRAISING STUDENT ACHIEVEMENT, SOME CLAIM THAT MULTIPLE CHOICE TESTS DEGENERATE INTO A GAME OF "BEAT THE MONKEY," I.E., ANSWERING MORE THAN 25 PERCENT OF THE ITEMS CORRECTLY, THE RATING EVEN A MONKEY COULD BE EXPECTED TO RECEIVE. SAMPLE STUDIES REVEAL THAT THE STUDENT WHO RAPIDLY SUPPLIES ANSWERS AT RANDOM MAY WELL PLACE HIGHER THAN THE SLOWER, SUPERIOR STUDENT. SUCH EXAMINATIONS AS THE N.Y. STATE REGENTS EXAMINATION IN RUSSIAN FOR SECONDARY SCHOOLS, THE COLLEGE ENTRANCE EXAMINATION BOARD'S RUSSIAN ACHIEVEMENT TEST, AND THE MLA COOPERATIVE FOREIGN LANGUAGE TESTS IN RUSSIAN ALL RELY HEAVILY, IF NOT COMPLETELY, ON MULTIPLE CHOICE. LANGUAGE EXAMINATIONS IN GENERAL COULD BE IMPROVED BY USING MORE SECTIONS SIMILAR TO THE MLA WRITING SECTION AND BY INCLUDING ESSAY QUESTIONS, ESPECIALLY IN STRUCTURED FORM, WHILE AN ACHIEVEMENT EXAMINATION MIGHT WELL REQUIRE INTERLINEAR CORRECTION OF BADLY GARBLED WRITING. FINALLY, FOREIGN LANGUAGE TEACHERS SHOULD COOPERATE WITH COLLEAGUES OF OUTSTANDING ABILITY IN OTHER FIELDS TO MINIMIZE THE ABUSES OF MULTIPLE CHOICE TESTS AND TO ENCOURAGE THE USE OF EXAMINATIONS REQUIRING STUDENTS TO ORGANIZE THEIR OWN THOUGHTS. THIS SPEECH WAS DELIVERED AT A MEETING OF THE NEW YORK AND NEW JERSEY GROUP OF THE AMERICAN ASSOCIATION OF TEACHERS OF SLAVIC AND EAST EUROPEAN LANGUAGES, PRINCETON, OCTOBER 29, 1966. (GJ)

ED013019

EXAMINATIONS - THEIR USE AND ABUSE

Pearl S. Bennett

A statement made in the Encyclopedia of Educational Research and verified in the dictionary is that there is no important distinction made between the terms tests and examinations in ordinary speech in the United States.(2) Since this is to be a rather ordinary speech, let us begin by examining the examinations in our extra-curricular lives and their usefulness.

Legend has it (and this one is not attributed to Confucius) that the Asian who inhabited the southern province of Canton would claim to be a thrice-a-day bather, and he would scornfully view his Northern Mandarin neighbor who supposedly bathed three times a year. However, the least bathed and most scorned would be the Mongolian, who allegedly would be bathed three times in his lifetime - at birth, at marriage, and at death. This can be related to physical examinations, at least for those of us who do not visit the medical doctor's office as frequently as we should. The body is subjected to a thorough physical examination at birth and at death, and, at least to the extent of a blood test, at marriage. Those patients who follow the rules will usually submit to yearly examinations for cancer and tuberculosis, and may even visit their dentist twice a year for an examination of their teeth and gums.

In their days of scouting, most boys and girls will

FL 000 118

pass a series of tests to get their merit badges. The ham-radio operator cannot operate without his license, earned by virtue of a test. Nor should the motor vehicle operator be caught without his driver's license, obtained, in the state of New Jersey, after passing a rigorous written examination, a behind-the-wheel test with an inspector, and a visual acuity test more perceptive than that given in most other states. (I have a cousin, a resident of New Jersey, who is color-blind. He cannot pass the New Jersey test for color blindness. He drives with a New York state license.)

When an actor or a musician wishes a job with a company, he takes an audition. Would not his audition be considered a type of examination?

The federal government gives examinations for the purposes of military service, civil service appointments and the Peace Corps. State boards conduct examinations in medicine and law, and for the applicants who would be certified as public accountants and professional engineers.

Most modern businesses confront their job applicants with placement tests in typing and shorthand, and for the use of office equipment and business machines.

We are still in the realm of extra-curricular testing, but drawing nearer to the source, when we consider the instruments which are intended to measure maximum capability, such a aptitude tests and intelligence tests. There are instruments which are intended to describe typical behavior, such as personality, interest and attitude inventories. There are also questionnaires and schedules used to rate

personal and social adjustment, and even sexual activities (remember the Kinsey reports?).

But let us turn our attention to a consideration of examinations in our scholarly lives. In educational usage, the term ordinarily refers to a series of questions or tasks designed to measure the knowledge or skill of an individual. Examinations are not only useful - they are necessary. Effective evaluation of student achievement with respect to accepted goals of instruction is considered an indispensable aspect of good teaching. The evaluation procedures that are used become a part of the instructional process and influence students in many ways.(2) They enable students to determine how well they are achieving. There is also the motivating effect of the knowledge of one's progress.

Educators on the secondary school level have found that examinations are useful for obtaining the following objectives: 1. to examine the validity of test questions

2. to evaluate students' retention of material over a short or comparatively long period of time

3. to evaluate the effectiveness of instruction and the effectiveness of particular teaching procedures and learning environments

4. to stimulate daily class work

5. to render educational guidance

6. to reveal areas of common difficulties or weaknesses as a basis for remedial work or re-teaching

7. to evaluate methods of selecting and organizing course materials

8. to improve the motivation of learning

9. to determine evidence of the amount of progress which students are making

10. to accumulate materials for research

Most educators are agreed that tests are an integral part of any formal program of education. While we are agreed as to the uses of examinations, there may very well be some disagreement as to the abuses of examinations. I shall first deal with a criticism of examinations in general, and then consider this abuse in examinations in the Russian language.

There are critics who feel that the multimillion-dollar testing business is approaching the proportions of a racket and an educational scandal.(1) They feel that the abuses of our examination system far outweigh the usefulness of examinations. One of the most dedicated of these Don Quixotes is Banesh Hoffman. Banesh Hoffman is a distinguished scientist and a long-time critic of multiple-choice tests. He gives a fully documented and reasoned account of the inadequacies and dangers of mechanical testing in his book <u>The Tyranny of Testing</u>.(3)

Have you complained lately that some of the ablest students are the least well prepared? Your observation may well have its source in the neglect of effort which multiple-choice testing entails. A student does not really know what he has learned until he has organized his material and explained it to someone else. The mere recognition of what is right in someone else's wording is only the beginning of the

awareness of truth.(3)

The essay test requires the student to demonstrate an important educational achievement - skill in written expression - which is not required by the objective test. As an educator, are not you concerned with the deleterious effects on teaching methods and on the curriculum generally which the use of multiple-choice tests alone might have?

One of the many cogent arguments against the multiple-choice test is the factor of chance guessing. Research has proven that candidates respond to multiple-choice test questions through 1. direct knowledge, 2. test-wiseness, 3. response sets and 4. chance guessing.(2)

"Do you want to play 'Beat the Monkey'?" That is the way a teacher of the "new" physics introduces the standardized multiple-choice tests to her students. What is 'Beat the Monkey'? On a multiple-choice test of one hundred items with four choices each, the laws of probability indicate that by random guessing, alone, even a monkey would get a score of twenty-five items correct.

To test the 'Beat the Monkey' theory for myself, I had thirty-two subjects number a paper from one to forty, and then letter at random, using A B C D for the odd numbers and F G H J for the even numbers. Then, I scored these "tests" with the hand-scoring key of the Russian Listening test, form MB, of the MLA Cooperative Foreign Language Tests. There are forty items, therefore the laws of probability indicate that with chance, alone, the subjects would score ten correct answers. Even a monkey could be expected to get ten

out of forty right. Twenty of the group of thirty-two were able to 'Beat the Monkey', that is, have more than ten correct answers. Two had exactly ten correct answers. Ten were not able to 'Beat the Monkey', having scores ranging from six to nine correct.

A set of thirty-two is hardly a scientific sampling, and I do not present it to you as such. I do not have statistical charts concerning the factor of chance guessing in any specific multiple-choice test. But I have made some simple observations from the point of view of an ordinary classroom teacher.

The scene is the quiet room where the Reading section of the MLA Cooperative Foreign Language Test in Russian is being administered. Student A is sitting behind student B. Student A has been a fine student, the best one in the class all year long. He has done his homework consistently and thoroughly. He has read many Russian books on his own. He takes the full thirty-five minutes to cover the Reading section, carefully studying each item, and writing the letter for his answer (A B C D F G H J) in the space provided. Student B is a good student, but not the best! He generally does his homework, but he does what is required of him, and nothing more. His class work has not been profound. Student B takes his test with one eye on the clock. Two minutes before the end of the test, I notice an odd behavior on student B's part. He is rapidly supplying letters to the blank spaces. He is supplying answers without any thought or knowledge of the Russian skills being tested. If only I

could be certain at what exact point student B started his random guessing, I could disqualify the ersatz answers. But I am not certain, and, in addition, as I look around the room in this last minute before I give the signal "Time's up" I see that other hands have started moving in a last minute flurry to supply a letter for every answer space, irregardless of the meaning of the Russian language. The most tragic feature of this test is that after grading the answers, I find that student A has thirty-three correct answers and ranks at the eighty-first percentile, while student B has guessed himself to forty-one correct answers and ranks in the ninety-first percentile.

At this point I should imagine that you are a)restless b)bored c)annoyed d)none of the above e)all of the above. I shall try not to bore you for much longer, just a few more minutes. In the mean time you can a)read your newspaper b) doze off c)scrutinize your neighbor d)all of the above.

I ask you to consider three different types of achievement examinations in the Russian language. Number one is the sample examination in Russian III, released by the Bureau of Foreign Languages Education, Albany (a "regents" examination for the secondary schools of New York state. Out of a total of one hundred possible points, eighty-five points are earned by virtue of multiple-choice items. This leaves fifteen points to be earned in two different sections. One section purports to be a part of the examination which is designed to test auditory comprehension, but, indeed, requires the student to write a grammatically complete answer.

The last section of the "regents" requires the candidate to write a letter in ten grammatically complete sentences for ten points.

The second type of examination I ask you to consider is a sample of the College Entrance Examination Board's Russian Achievement Test as it appeared in one of those cram books that the students can buy from anywhere from $2.98 and up, and might have the picturesque title of, for example, "How to Score Higher than Your Next Door Neighbor on Your College Entrance Exams So That He'll Be Sure to Invite You to the Senior Prom" subtitle "Pass High and Win the Boy of Your Dreams!". This test has one hundred items. All one hundred of the items are multiple-choice.

The third type of examination to be considered is the MLA Cooperative Foreign Language Tests in Russian. In December 1964 at our AATSEEL convention in New York City, one of our colleagues presented an analysis of the MLA tests. I believe that the Writing sections of the MLA tests are excellent, although the method of scoring is sometimes inadequate. The scale of three, two, one or zero points provides for most accurate scoring of a three part item, however, some questions in one set contain either more or less than three elements, so that a fair scoring becomes a matter of personal indiscretion.

The Listening section of the MLA tests is misnamed. Only the first four items of form LA and LB merit the label of "listening". For the first four items, the student hears a sentence and selects his answer from a series of four pic-

tures. Items #5 through #45 of LA and LB and all forty items of MA and MB Listening are, in truth, Listening and Reading items. The candidate hears a statement, but then he must read the four choices for himself before he can select the correct response. What does the student do? The voice on the tape has told him what to do. "It will be to your advantage to answer every question even though you may not be sure that your answer is correct." So, the candidate answers every question. He plays 'Beat the Monkey' and he walks out of the examination room wondering why he has been working so hard to master the Russian language. The Listening section takes approximately twenty-five minutes, and the Reading section of fifty multiple-choice items takes thirty-five minutes. More than half of the candidate's time is spent playing games.

Although the title of this paper limits our concern to the uses and abuses of examinations, I would like to follow through with a few suggestions. In the first place, I would like to see more examinations similar to the Writing section of the MLA tests. Secondly, I would like to see examinations which use essay questions. The essay questions can be highly structured. Research has shown that scoring can be quite reliable with reference to the facts the student should recall and mention when essay exercises are structured. Thirdly, I would like to see an interlinear test on an achievement examination. The interlinear test presents the candidate with a triple-spaced copy of a badly garbled piece of writing. The candidate is allowed a specified a-

mount of time to indicate necessary corrections and deletions, but is instructed not to add ideas of his own. The candidate's paper is then scored for his treatment of predetermined errors in the copy.(2)

Finally, we must join forces with our colleagues in other disciplines in order to do something about the abuses of multiple-choice examinations. Could we join a committee of inquiry whose minimum concern would be the quality of multiple-choice tests and their manufacturers? The committee should include creative people of commanding intellectual stature who would bring fresh vision to the testing situation.(3) The committee would realize how important it is to train students to organize their own thoughts and to put something of themselves into a project, and how damaging it can be to reward students for merely picking wanted answers at rates of up to one hundred an hour.

Dr. Frederick M. Raubinger is quoted as having said, "Tests alone cannot substitute for the wise and mature judgment of those who know children intimately as human beings and who refuse to regard children in terms of a series of data recorded on an IBM punch card."(1)

Perhaps there is hope. Perhaps each one of us here can perform a small part, and eventually we shall succeed in ridding the educational scene of the insulting-to-the-intelligence abuses of contemporary multiple-choice examinations.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

REFERENCES

1. Fine, Benjamin. The Modern Family Guide to Education. Doubleday and Company, Inc. 1962. 648 p.

2. Harris, Chester W. ed. Encyclopedia of Educational Research, 3rd. edition. Macmillan Company. 1960. 1564 p.

3. Hoffman, Banesh. The Tyranny of Testing. The Crowell-Collier Press. 1962. 223 p.